An Age-Length Key for Sockeye Salmon Smolt

By:

Brian G. Bue

and

Douglas M. Eggers

Regional Information Report[1] No. 2D89-5

Alaska Department of Fish and Game
Division of Commercial Fisheries, Central Region
333 Raspberry Road
Anchorage, Alaska 99518

April 1989

[1]Contribution 89-5 in the Bristol Bay Area Data Report Series. The Regional Information Report Series was established in 1987 to provide an information access system for all unpublished Divisional reports. These report frequently serve diverse ad hoc informational purposes or archive basic uninterpreted data. To accommodate needs for up-to-date information, reports in this series may contain preliminary data.

AUTHORS

Brian G. Bue is the Region II Stock ID Biometrician for the Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage 99518.

Douglas M. Eggers is the Chief Fisheries Scientist for the Alaska Department of Fish and Game, Division of Commercial Fisheries, P.O. Box 3-200, Juneau 99802.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

An age-length key was developed to assign ages to sockeye salmon (*Oncorhynchus nerka*) smolt based on smolt length. Age-weight-length data were obtained from a subsample of all measured fish to estimate the age-length key. The key was then used to assign ages to the rest of the sample which were measured for length only. This paper documents the algorithm used as well as presents a cursory analysis of its performance.

KEYWORDS: Age-Length Key, juvenile sockeye salmon, *Oncorhynchus nerka*

INTRODUCTION

Numbers by age of seaward migrating sockeye salmon (*Oncorhynchus nerka*) smolt are estimated for 5 of the major river systems in Bristol Bay (Bue et al. 1988). Age composition can change dramatically through time in a single season. It is not uncommon to observe 100% age-2 smolt early in the run, 100% age-1 smolt late, and approximately an even split in age composition at some time between.

Prior to 1983, 120 smolt were sampled daily. It was determined that this level of sampling did not provide the desired level of precision and accuracy (within 5% of the actual percentage, 95% of the time) on a daily basis; consequently, sampling goals were increased to 400 smolt per day in 1983 and further increased to 6 samples of 100 smolt per day in 1986 to account for observed clustering by age group.

Collection of standard age-weight-length (AWL) data is a very time consuming process. Fork length (mm) and weight (g) are measured while age is determined from scale smears mounted on glass slides and later read using a microfiche reader. An age-length key was developed to reduce the time required to age 400 smolt each day. This allowed subsampling for AWL information (usually 100 smolt per day) with the remaining fish measured for length only (300 or more per day). The age-length key was then estimated from AWL samples and used to classify the remaining smolt lengths into age groups.

This paper documents the age-length key and presents a cursory sensitivity analysis of the procedure.

Age-Length Key

The objective of the age-length key is to categorize smolt as either age-1 or -2 based on fork length. This is done by determining the critical length (L^*) which minimizes classification error (E_i, where i=age).

E_1=Number of smolt classified as age-2 given they are age-1 , and
E_2=Number of smolt classified as age-1 given they are age-2 .

L^* is chosen such that $E_1 = E_2$. (1)

A prior estimate of age composition can be made based on the proportions of age-1 and -2 smolt (p_i) in the AWL sample. These proportions multiplied by the total population estimate give the estimated number of smolt by age.

It is assumed smolt length (L) by age is normally distributed about mean (μ_i) with standard deviation (σ_i) and that $\mu_1 < \mu_2$ (Figure 1). Thus the probability that a fish of length L and known age-1 (L_1) is misclassified as age-2 is;

$$P(L_1 > L^*) = 1 - F\left(\frac{L^* - \mu_1}{\sigma_1}\right) \quad , \qquad (3)$$

and the probability that a fish of length L and known age-2 (L_2)is misclassified as age-1 is;

$$P(L_2<L^*) = F(\frac{L^*-\mu_2}{\sigma_2}) \quad , \tag{2}$$

where F is the normal distribution function.

The classification errors can now be expressed as;

$$E_1 = Np_1 \, [1 - F(\frac{L^*-\mu_1}{\sigma_1})] \qquad , \text{ and} \tag{4}$$

$$E_2 = Np_2 \, F(\frac{L^*-\mu_2}{\sigma_2}) \quad . \tag{5}$$

Substituting for E1 and E2 in expression 1 gives;

$$Np_1 \, [1 - F(\frac{L^*-\mu_1}{\sigma_1})] = Np_2 \, F(\frac{L^*-\mu_2}{\sigma_2}) \quad . \tag{6}$$

This equality can be rearranged such that;

if $p_1 \geq p_2$

$$L^* = \mu_1 + \sigma_1 F^{-1}[1 - \frac{p_2}{p_1} F(\frac{L^*-\mu_2}{\sigma_2})] \qquad , \text{ or} \tag{7}$$

if $p_1 < p_2$

$$L^* = \mu_2 + \sigma_2 F^{-1}[\frac{p_1}{p_2} (1 - F(\frac{L^*-\mu_1}{\sigma_1})] \quad . \tag{8}$$

The equality is arranged conditional on the magnitudes of p1 and p2 due to the constraints of F^{-1} (F^{-1} is only defined between 0 and 1). The sample estimates for μ_i and σ_i ($\bar{x}_i$ and s_i) are substituted into equation 7 or 8 to solve for L^*.

L^* is solved for iteratively by selecting a starting value; generally,

$$L^*_{est} = \frac{\bar{x}_1 - \bar{x}_2}{2} \quad . \tag{9}$$

L^*_{est} is substituted into the right side of equation 7 or 8 which is then evaluated.

$$\text{If } |L^* - L^*_{est}| = 0.0 \ , \tag{10}$$

the solution has been found.

$$\text{If } |L^* - L^*_{est}| \neq 0.0 \ , \tag{11}$$

a new L^*_{est} is estimated;

$$L^*_{est} = \frac{L^* + L^*_{est}}{2} \quad , \tag{12}$$

and substituted into equation 7 or 8 and evaluated once again. The iterations continue until expression 10 is satisfied.

MATERIALS AND METHODS

Performance of the age-length key was examined by simulating age-length samples of varying proportions and standard deviations. While the number of possible combinations of mean lengths, age proportions, and standard deviations is limitless, only 2 cases were examined: 1) age-1 and -2 populations with means of 85.0 mm and 95.0 mm, respectively, and standard deviation of 3.0 for both groups, and 2) age-1 population with mean of 85.0 mm standard deviation of 4.0 and age-2 population with mean of 95.0 standard deviation 7.0. A wide range of age proportions were examined for each case (age-1/age-2: 0.95/0.05, 0.90/0.10, 0.80/0.20, 0.60/0.40, 0.50/0.50, 0.40/0.60, 0.20/0.80, 0.10/0.90, 0.05/0.95). The means, standard deviations, and proportions were within the range typically observed.

Two age-length groups for each case and proportion combination were generated from normal distributions with set mean and standard deviation using the IMSL (1987) random number generator (IMSL routine DRNNOF). One group was used to estimate an age-length key (standard group), the other to evaluate the key's accuracy (validation group). The software used to estimate the age-length key made use of the IMSL (1987) normal and inverse normal distribution subroutines (IMSL routines DNORDF and DNORIN)[Appendix A].

RESULTS AND DISCUSSION

The age-length key performed well. Estimated proportions differed from known proportions by less than 1% in 10 of the 18 simulations and by less than 2% in 16 of the 18 (Tables 1 and 2). The greatest difference was 3.2% in the unequal standard deviation case. The case with equal standard deviations seemed to do slightly better (NSC).

This technique for assigning ages based on lengths appears to work reasonably well. The range of age proportions examined resemble those actually observed while only rarely do standard deviations differ between ages as much as those used in the second case. This procedure was not examined for robustness with regards to deviations from the assumption of normality.

LITERATURE CITED

Bue, B.G., D.L. Bill, W.A. Bucher, S.M. Fried, H.J. Yuen, and R.E. Minard. 1988. Bristol Bay sockeye salmon smolt studies for 1986. Alaska Department of Fish and Game, Division of Commercial Fisheries, Technical Fishery Report No. 88-15, Juneau.

IMSL. 1987. Stat/Library - Fortran subroutines for statistical analysis, version 1.0. Houston, Texas.

Table 1. Performance of the age-length key on age-length groupings with the same standard deviation but varying proportions.

	Standard Group				Validation Group				Proportion		
Age[a]	$\bar{x}$	s	n	L^*	$\bar{x}$	s	n	E_i	Known	Estimate	|Diff.|[b]
1	84.91	2.86	500	89.97	85.01	3.08	500	35	.500	.489	.011
2	95.09	2.89	500		94.87	2.98	500	24	.500	.511	.011
1	85.04	3.21	600	90.63	85.24	2.91	600	19	.600	.604	.004
2	95.16	2.93	400		95.06	2.71	400	23	.400	.396	.004
1	85.25	3.10	400	89.69	84.95	2.99	400	25	.400	.400	.000
2	94.76	3.10	600		95.01	3.05	600	25	.600	.400	.000
1	84.90	3.04	800	90.71	85.07	2.96	800	26	.800	.786	.014
2	94.46	3.08	200		95.37	2.93	200	12	.200	.214	.014
1	85.23	3.10	200	88.97	85.00	3.22	200	28	.200	.184	.016
2	94.99	3.16	800		94.90	2.90	800	12	.800	.816	.016
1	85.19	2.98	900	91.88	85.08	2.96	900	11	.900	.894	.006
2	95.48	2.94	100		95.59	3.00	100	5	.100	.106	.006
1	84.54	3.16	100	88.36	84.82	3.32	100	12	.100	.107	.007
2	95.08	3.00	900		94.86	3.01	900	19	.900	.893	.007
1	85.08	3.07	950	92.17	85.05	2.98	950	7	.950	.953	.003
2	94.79	3.13	50		95.15	3.13	50	10	.050	.047	.003
1	84.69	3.00	50	87.99	85.54	2.97	50	11	.050	.042	.008
2	95.07	2.88	950		94.97	3.00	950	3	.950	.958	.008

a Age-1 randomly selected from a normally distributed population with mean of 85.0 mm and standard deviation 3.0.

Age-2 randomly selected from a normally distributed population with mean of 95.0 mm and standard deviation 3.0.

b |Diff.| is the absolute difference between the known and estimated proportions.

Table 2. Performance of the age-length key on age-length groupings with different standard deviations and varying proportions.

	Standard Group				Validation Group				Proportion		
Age[a]	$\bar{x}$	s	n	L^*	$\bar{x}$	s	n	E_i	Known	Estimate	|Diff.|[b]
1	84.87	3.94	500	88.33	84.92	4.00	500	87	.500	.489	.011
2	95.21	7.83	500		95.29	6.97	500	76	.500	.511	.011
1	84.69	3.91	600	89.27	84.72	4.11	600	65	.600	.606	.006
2	95.20	6.49	400		95.66	6.82	400	71	.400	.394	.006
1	84.91	3.90	400	87.89	85.05	4.16	400	115	.400	.368	.032
2	95.22	7.04	600		95.12	7.02	600	83	.600	.632	.032
1	84.77	3.83	800	90.40	85.05	4.04	800	71	.800	.781	.019
2	94.30	6.76	200		95.38	7.67	200	52	.200	.219	.019
1	84.80	3.78	200	85.87	84.82	3.81	200	88	.200	.204	.004
2	94.91	6.96	800		94.66	7.25	800	92	.800	.796	.004
1	85.06	3.93	900	91.96	84.97	3.93	900	43	.900	.885	.015
2	94.34	6.46	100		94.71	6.30	100	28	.100	.115	.015
1	84.81	3.72	100	84.73	84.89	3.69	100	48	.100	.105	.005
2	95.15	6.58	900		95.11	7.03	900	53	.900	.895	.005
1	84.93	4.02	950	92.90	85.21	4.10	950	38	.950	.925	.025
2	93.85	7.17	50		97.68	7.36	50	13	.050	.075	.025
1	85.37	3.88	50	82.28	84.61	3.76	50	34	.050	.052	.002
2	94.73	7.18	950		94.81	7.00	950	36	.950	.948	.002

[a] Age-1 randomly selected from a normally distributed population with mean of 85.0 mm and standard deviation 4.0.

Age-2 randomly selected from a normally distributed population with mean of 95.0 mm and standard deviation 7.0.

[b] |Diff.| is the absolute difference between the known and estimated proportions.

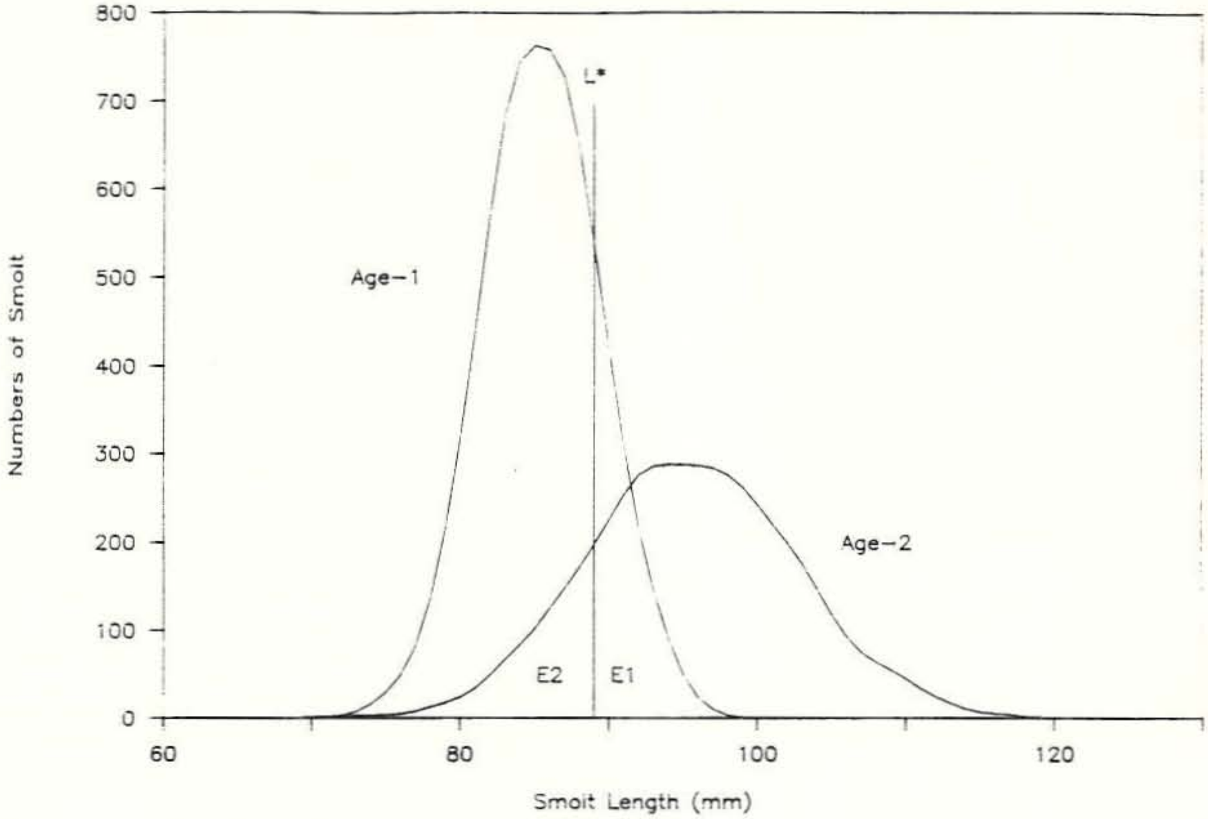

Figure 1. Typical distribution of smolt lengths by age.

Appendix A. Fortran source code for the age-length key.

```
$DEBUG
C...PROGRAM AGELNK
C...THIS PROGRAM FINDS OPTIMAL CLASSIFICATION
C...FOR AGE OF SMOLT BASED ON LENGTH
C...TWO AGE GROUPS (ONE CHECK AND TWO CHECK) OF
C...SMOLT ARE CONSIDERED
      IMPLICIT DOUBLE PRECISION (A-H)
      IMPLICIT DOUBLE PRECISION (O-Z)
      EXTERNAL DNORDF, DNORIN
      WRITE(*,1)
1     FORMAT(' ')
      WRITE(*,2)
2     FORMAT('  This program finds the optimal classification',
     &/,'  for age of smolt based on length. ',
     &/,'  Version 2.0 (Bue 3/89)  ')
      WRITE(*,1)
      WRITE(*,3)
3     FORMAT(' ','MEAN LENGTH, SD FOR AGE I,  (XM1,SD1)')
      READ(*,4) XM1,SD1
      WRITE(*,1)
      WRITE(*,5)
5     FORMAT(' ','MEAN LENGTH, SD FOR AGE II, (XM2,SD2)')
      READ(*,4) XM2,SD2
      WRITE(*,1)
      WRITE(*,6)
6     FORMAT(' ','PROPORTION OF AGE I, AGEII,  (P1,P2)')
      READ(*,4) P1,P2
4     FORMAT(2F10.0)
      WRITE(*,1)
      WRITE(*,1)

C...ENTER INTERATIVE EQUATION SOLVER
C...DNORDF IS IMSL DBL PRECISION NORM. DIST. FUNCTION
C...DNORIN IS IMSL DBL PRECISION INVERSE NORM. DIST. FUNCTION

      IT=1
      XX=(XM1+XM2)/2.
      IF(P1.LT.P2) GO TO 200

100   S2=(XX-XM2)/SD2
      Z1=1-((P2/P1)*DNORDF(S2))
      XN=XM1+(SD1*DNORIN(Z1))
      ER=ABS(XX-XN)
      IF(ER.LT.0.01) GO TO 300
      WRITE(*,110) IT,XX,XN
110   FORMAT(' Iteration ',I3,2X,'L =',F8.2,2X,'Lest =',F8.2)
      XX=(XX+XN)/2.
      IT=IT+1
      GO TO 100
```

```
200   S1=(XX-XM1)/SD1
      Z1=(P1/P2)*(1-DNORDF(S1))
      XN=XM2+(SD2*DNORIN(Z1))
      ER=ABS(XX-XN)
      IF(ER.LT.0.01) GO TO 300
      WRITE(*,110) IT,XX,XN
      XX=(XX+XN)/2.
      IT=IT+1
      GO TO 200

300   CONTINUE
      S1=(XN-XM1)/SD1
      S2=(XN-XM2)/SD2
      WRITE(*,1)
      WRITE(*,310) IT,XN
310   FORMAT(' No. Iterations =',I3,4X,'Critical Length (L*) =',F8.2)
      WRITE(*,1)
      STOP
      END
```